Impacts of Targeted AERI and Doppler Lidar Wind Retrievals on Short-Term Forecasts of the Initiation and Early Evolution of Thunderstorms

MICHAEL C. CONIGLIO

NOAA/National Severe Storms Laboratory, Norman, Oklahoma

GLEN S. ROMINE

National Center Atmospheric Research, Boulder, Colorado

DAVID D. TURNER

NOAA/Earth System Research Laboratory, Boulder, Colorado

RYAN D. TORN

University at Albany, State University of New York, Albany, New York

ABSTRACT

The ability of Atmospheric Emitted Radiance Interferometer (AERI) and Doppler lidar (DL) wind profile observations to impact short-term forecasts of convection is explored by assimilating retrievals into a partially cycled convection-allowing ensemble analysis and forecast system. AERI and DL retrievals were obtained over 12 days using a mobile platform that was deployed in the preconvective and near-storm environments of thunderstorms during the afternoon in the U.S. Great Plains. The observation locations were guided by real-time ensemble sensitivity analysis (ESA) fields. AERI retrievals of temperature and dewpoint and DL retrievals of the horizontal wind components were assimilated into a control experiment that only assimilated conventional observations. Using the fractions skill score within 25-km neighborhoods, it is found that the assimilation of the AERI and DL retrievals results in far more times when the forecasts are improved than degraded in the 6-h forecast period. However, statistical confidence in the improvements often is not high and little to no relationships between the ESA fields and the actual changes in spread and skill is found. But, the focus on convective initiation and early convective evolution—a challenging forecast problem—and the fact that frequent improvements were seen despite observations from only one system over a limited period, provides encouragement to continue exploring the benefits of ground-based profilers to supplement the current upper-air observing system for severe weather forecasting applications.

1. Introduction

Our ability to measure the atmospheric state above the ground needs to be improved in order to meet the growing needs of society (National Research Council 2009; Stalker et al. 2013). One way to address these needs is through ground-based remote sensing systems that can fill in large spatial and temporal gaps in the current upper-air network (Geerts et al. 2018). Ground-based profilers present an intriguing option because of the high time resolution of their retrievals and because

they can be left unattended for long periods of time. To explore the use of these systems for severe weather applications, the National Severe Storms Laboratory obtained an Atmospheric Emitted Radiance Interferometer (AERI) to retrieve profiles of temperature and water vapor and a Doppler lidar (DL) to retrieve wind profiles. These profilers are mounted on a trailer—named the Collaborative Lower Atmosphere Mobile Profiling System (CLAMPS)—to allow for targeted observations (Wagner et al. 2019).

Characteristics of severe convective weather are sensitive to the state of the convective boundary layer (CBL). Unfortunately numerical weather prediction (NWP) of the CBL state continues to be error prone

Corresponding author: Michael C. Coniglio, Michael.Coniglio@noaa.gov

with systematic biases (Coniglio et al. 2013; Cohen et al. 2015; Shin and Dudhia 2016; Cohen et al. 2017). Many efforts to reduce these biases on convection-allowing (~1–4 km) grids through improved physical parameterizations are ongoing (e.g., Shin and Hong 2015; Benjamin et al. 2016; Zhou et al. 2018). In the meantime, although a biased background state can hinder the reduction of analysis error through the assimilation of observations (Romine et al. 2013), the upper-air network is exceedingly sparse relative to convective space and time scales, and therefore it is likely that even biased systems would still benefit from additional, judiciously chosen observations. Indeed, despite the presence of known temperature and humidity biases in the CBL, some improvements to short-term (~1–6 h) convection-allowing forecasts of thunderstorms has been shown by assimilating special upper-air observations from radiosondes (Hitchcock et al. 2016; Coniglio et al. 2016).

For NWP applications, it is important to gauge the worth of new observing systems through multiple diagnostic studies of forecast-system improvement (National Research Council 2009). To contribute to this effort, we have assimilated retrievals from the CLAMPS AERI and DL into an ensemble analysis and forecast system that has been used for real-time prediction of convective weather. This study is inspired by the Mesoscale Predictability Experiment (MPEX; Weisman et al. 2015) in which ensemble sensitivity analysis (ESA; Ancell and Hakim 2007; Torn and Hakim 2008) was used to guide where and when to take observations to increase the likelihood of making an impact on forecasts. In MPEX, a Gulfstream V aircraft released dozens of dropsondes in the early morning over the Intermountain West with a focus on gauging impacts on 12–24-h forecasts (Romine et al. 2016). In this study, CLAMPS is used to obtain AERI and DL observations from a single location and we focus on 1–6-h forecast impacts. It is shown here that retrievals from the AERI and DL, even from only one location, can have a positive impact on analyses of the prestorm environment and the subsequent prediction of storms. Section 2 provides an overview of the CLAMPS AERI and DL and the methods to obtain retrievals of temperature, humidity, and winds. Section 3 describes the ESA method used for observation targeting. Section 4 describes the ensemble analysis and forecasting system, the methods of postprocessing of the retrievals for data assimilation, and the forecast verification methods. Pertinent results from the data assimilation experiments are summarized in section 5, experiment sensitivities are discussed in section 6, and conclusions are provided in section 7.

2. Ground-based profilers

a. AERI

The AERI is a passive remote sensor that measures downwelling spectral infrared radiation every 20 s in a portion of the spectrum (3.3–19.2 μm) that is sensitive to the vertical thermodynamic structure of the atmosphere (Knuteson et al. 2004). After applying a principal component–based noise filter to reduce random error in the radiance spectra (Turner et al. 2006), and averaging the radiances to 2-min intervals, the radiances are processed through an optimal-estimation-based retrieval algorithm (AERIoe) described in Turner and Löhnert (2014). AERIoe is an iterative Guass–Newton retrieval technique that uses the Clough and Iacono (1995) radiative transfer model to obtain estimates of the vertical profile of temperature t and water vapor mixing ratio q, as well as the cloud liquid water path (LWP) and mean cloud effective radius in the column. The retrieval is constrained in the middle to upper troposphere by a first guess from the Rapid Refresh model analyses (Benjamin et al. 2016), but the final retrievals are insensitive to the particular first-guess profile that is used (Turner and Löhnert 2014). More information on AERIoe can be found in Turner and Löhnert (2014) and Turner and Blumberg (2018).

The retrieved t and q profiles lose vertical resolution rapidly with height [as shown later and in Turner and Löhnert (2014)], which reflects sensitivities in the radiative transfer model, truncation error in applying the radiative transfer model to a discrete set of height points, level-to-level covariance in the a priori data used to constrain the retrieval, and noise in the observed radiance spectra. Because of these uncertainties, the retrievals contain far fewer independent data points than what can be obtained from in situ methods (e.g., radiosondes). However, assimilation of these above-ground observations of t and q every 2 min, despite their low vertical resolution, may be an effective way to incorporate mesoscale environmental heterogeneity (if present) into model initial conditions and/or reduce initial condition errors. Further, the retrievals include estimates of their associated error as a function of height (Turner and Löhnert 2014), which is important to quantify for data assimilation purposes.

b. Doppler lidar

Doppler lidar (light detection and ranging) is a visible analog to Doppler radar that uses light backscattered from aerosols to measure line-of-sight particle velocity. NSSL uses a pulsed Doppler lidar manufactured by Halo Photonics Ltd. and is similar to the lidar described in Pearson et al. (2009) (see Table 1 for the primary lidar parameters used in this work). For this study, the lidar

TABLE 1. Parameters of the lidar.

Wavelength	1.5 μm
Pulse repetition frequency	15 kHz
Nyquist velocity (bandwidth)	19.4 m s^{-1}
Sampling frequency	30 MHz
Points per range gate	6
Pulse length	18 m
Range gate length	30 m
Averaging time per ray	1 s
Minimum range	90 m
Maximum range	12 000 m

was configured to complete a full conical scan at 60° elevation in 8 steps (every 45° in azimuth). A full conical scan takes ~20 s to complete and these scans were performed every three minutes (the lidar was scanning vertically when not doing conical scans). In the ideal case of a homogeneous wind field, the line-of-sight winds are sine-like as a function of azimuth and the wind components are retrieved through fitting a sine function to the radial wind speed (commonly known as the velocity–azimuth display, or VAD, technique). A separate sine wave is fit to the winds at each ~30-m range gate to obtain a high-resolution vertical profile of the wind every three minutes. The vertical range of the DL sample is limited by the amount of aerosol backscatter and clouds (the light can penetrate only very thin clouds). For the plains region sampled, the aerosol concentration can be large enough to allow for retrievals up to 3 km AGL but a retrieval depth of 1–1.5 km is typical.

3. Targeting using ensemble sensitivity analysis

The AERI and DL observations used in this study were obtained from 12 CLAMPS deployments made in 2016–17 in the U.S. Great Plains (Fig. 1). When evaluating impacts of observation systems, particularly when observing at only one location as in this study, it is important to attempt to gauge the forecast sensitivity to that observation since a lack of forecast impact could be related to a lack of environmental sensitivity at the sampled location and not the ability of the observing system to positively impact forecasts. Like in MPEX, guidance for where to deploy CLAMPS in the preconvective environment was provided by ESA fields that were derived from a 1200 UTC initialized 30-member version of the National Center for Atmospheric Research (NCAR) real-time ensemble (Schwartz et al. 2015). In general, the sensitivity of the expected value (e.g., the ensemble mean) of a forecast metric *J* to a state variable *x* at an earlier time is defined as
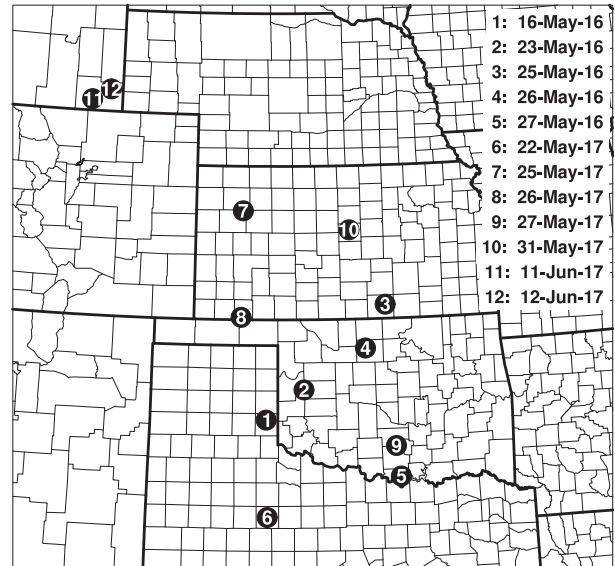

FIG. 1. Locations of the 12 CLAMPS deployments for which data are used for the data assimilation experiments. Durations of the AERI and DL retrievals from the deployments and the convective mode of the event are provided in Table 3.

$$\frac{\partial J}{\partial x} = \frac{\text{cov}(\mathbf{J}_i, \mathbf{x}_i)}{\text{var}(\mathbf{x}_i)}, \tag{1}$$

where $\mathbf{J}_i$ and $\mathbf{x}_i$ are the ensemble member estimates of *J* and *x*, respectively; cov is the covariance; and var is the variance (Torn and Hakim 2008). The sensitivity can be thought of as a linear regression of the ensemble forecast metric onto an earlier ensemble state variable (Torn and Hakim 2008) where the slope of the regression is the sensitivity.

Two types of sensitivity plots were used as guidance for observation targeting. First, Eq. (1) was computed at each grid point using 6- or 9-h forecasts to define the state variables [*x* in Eq. (1)].[1] In 2016, the forecast metric [*J* in Eq. (1)] was the simulated composite (column maximum) reflectivity averaged over a 2-h period within a ~250 km (east–west) by ~300 km (north–south) region. The location of this region was chosen to encompass forecasts of storms on that day and the 2-h period was chosen to end at the time that the standard deviation in vertical velocity within the region of forecasted storms

---

[1] Since real observation targeting takes time to deploy instruments, and there is always a latency in producing NWP model forecasts, model state variables (*x*) need to be evaluated from forecasts rather than initial conditions. In MPEX, given practical limitations of observation targeting with an aircraft, 24–30-h forecasts were used to define the state variables for 36-h forecast metrics. In this study, 6–9-h forecasts are used to define the state variables for 11–13-h forecast metrics.

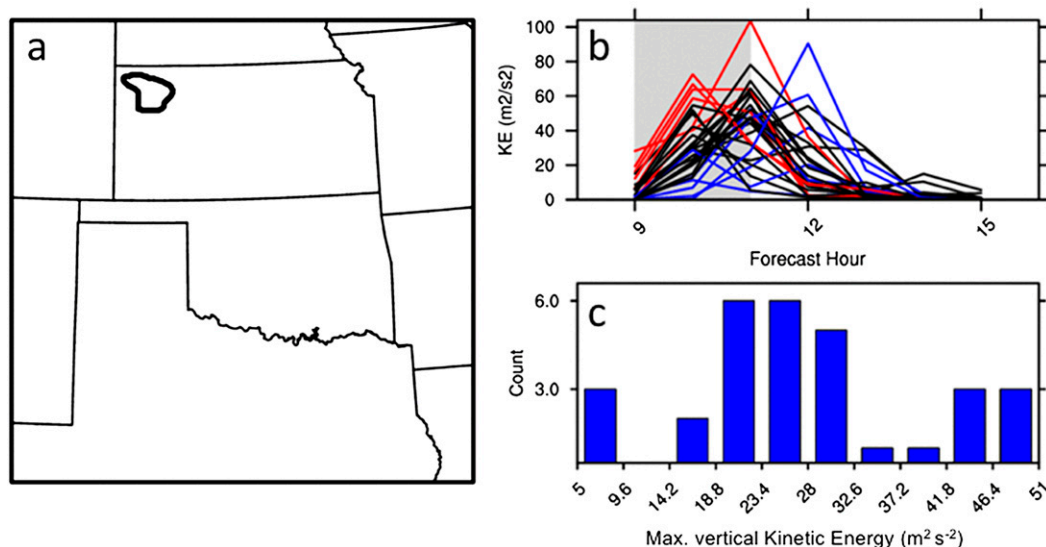## 2017052512 Max. Vertical Kinetic Energy Metric

FIG. 2. (a) The thick black contour encloses the forecast metric area that maximizes the variance in maximum vertical kinetic energy (MVKE) among the 30 members of the NCAR real-time ensemble averaged over 9–11-h forecasts (valid at 2100–2300 UTC 25 May 2017). (b) MVKE ($m^2 s^{-2}$) averaged over the area shown in (a) with the 10 members with the highest (lowest) averaged MVKE averaged in the 9–11-h period (gray shading) are shown in red (blue). (c) Histogram of the MVKE averaged in the area shown in (a) and over 9–11 h.

was maximized. This 2-h period was chosen to capture convective initiation and early convective evolution in the forecasts. In 2017, the forecast metric was changed to be the maximum vertical kinetic energy (MVKE) and a dynamic procedure was developed to find 2-h periods and areas that contain large variance of MVKE. To illustrate the procedure on a given day, on 25 May 2017 CLAMPS was deployed near an area of large sensitivity of MVKE averaged in the area shown in Fig. 2a to 6-h forecasts of potential temperature averaged in the lowest 1 km (Fig. 3a).[2] The sensitivity appeared to be related to uncertainty in temperature south of a cold front and east of a dryline (Fig. 3a). The positive sensitivity values near CLAMPS shown in Fig. 3a indicate that ensemble members with warmer CBLs at 1800 UTC resulted in higher values of MVKE averaged in the area shown in Fig. 2a between 2100 and 2300 UTC.

While real-time plots of Eq. (1) are helpful in identifying potentially sensitive areas, and provide a means to interpret physical reasons for differences in the forecasts (Torn and Romine 2015; Hill et al. 2016), quantitative estimates of how an observation will impact the later ensemble mean forecasts requires knowledge of the observation value itself, so that one could determine errors in the model state estimates. An observation will not change the ensemble mean forecasts if there is no error in the ensemble mean model state, regardless of the sensitivity. For real observation targeting, it can be difficult to determine ahead of time when and where the model state will have significant errors, and thus Eq. (1) can be difficult to exploit in real time.

To better quantify how an observation will impact forecasts, Ancell and Hakim (2007) derived an expression that estimates the reduction in the ensemble *variance* of J that would occur by assimilating an observation at a particular location, termed the observation impact value [see Eq. (6) in Torn and Hakim (2008)]. This technique estimates the reduction of forecast variance by only knowing observation error characteristics and ensemble forecast values, $\mathbf{J}_i$; the observation value itself is not required. A key assumption is that a reduction in ensemble forecast variance should translate into an increase in forecast skill (given a well-calibrated ensemble). An example of observation impact values that were
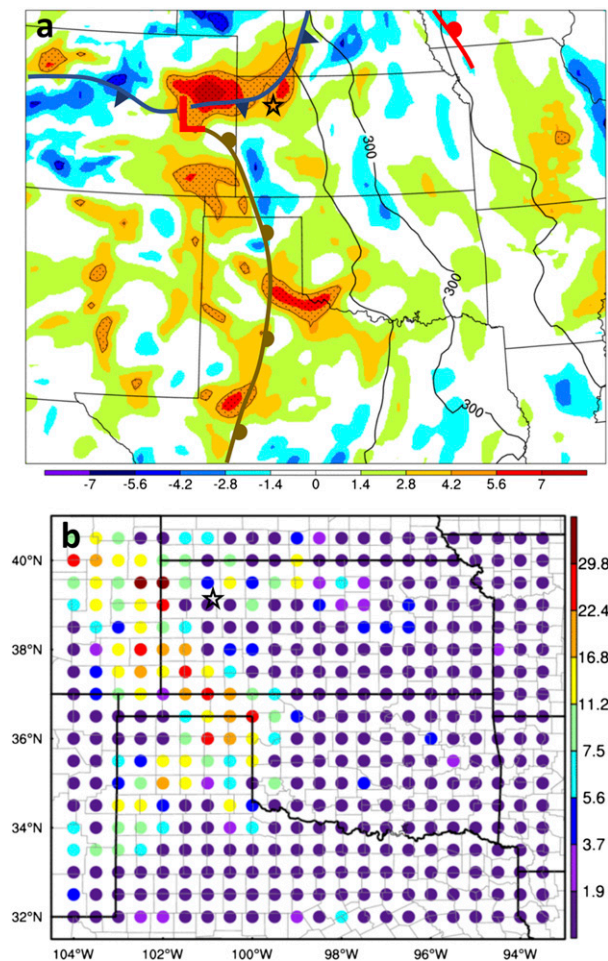
---

[2] This example illustrates a trade-off that was often made in deploying CLAMPS. While the maximum in sensitivity was located to the north and west of the CLAMPS deployment location, it was desirable to obtain as long of a time series of observations as possible prior to convective initiation that was also within the 6–9-h period in which state variables were used to compute the sensitivity fields. In this case, CLAMPS had a long ferry time that morning and a longer time series of observations prior to convective initiation was preferred over locating CLAMPS in the area with the highest sensitivity, which would have taken too much time.

FIG. 3. (a) The sensitivity of maximum vertical kinetic energy ($m^2 s^{-2}$) averaged over 9–11-h forecasts in the area shown in Fig. 2a ($\overline{MVKE}$) to the potential temperature averaged in the lowest 1 km ($\overline{\theta}$) in 6-h forecasts valid 1800 UTC 25 May 2017. Units are K^{-1} since the base sensitivity values of $m^2 s^{-2} K^{-1}$ are normalized by the ensemble standard deviation of $\overline{MVKE}$ within the metric area. Positive (negative) values indicate that higher values of $\overline{\theta}$ in 6-h forecasts at that location relate to more (less) $\overline{MVKE}$. Hatched areas show 95% statistical significance, that is, the absolute value of the regression coefficient is greater than its 95% confidence bounds computed from the ensemble data. CLAMPS was deployed at the location of the star starting at ~1800 UTC. The fronts and symbols are drawn manually from a surface analysis. (b) Observation impact values, or the hypothetical reduction in the variance of $\overline{MVKE}$ given the assimilation of t, t_d, u, and v below ~3 km AGL, valid at 1800 UTC. Front and dryline symbols are drawn manually from MADIS METAR and mesonet observations.

produced in real time is provided in Fig. 3b. The model state variable in this technique is a linear combination of t, t_d, and the horizontal wind components (u and v) over the lowest ~3 km AGL in 6-h forecasts. The values of each dot in Fig. 3b give the hypothetical reduction in the variance of 9–11-h MVKE averaged over the area shown in Fig. 2a if observed lower-tropospheric profiles

of t, t_d, u, and v were assimilated at the location of the dot at 1800 UTC.

It should be noted that areas identified as having the highest ensemble sensitivity for one variable are not necessarily the same for other variables, and are not necessarily the same areas where the possible variance reduction from assimilation of observations is maximized. This leads to some ambiguity in where to target for observations to maximize their impact. The observation impact values are directly proportional to the forecast-metric-observation covariance [see Eq. (6) in Torn and Hakim (2008)], meaning regions with large ensemble variance in reflectivity are more likely to have large observation impact values than regions with small forecast-metric-observation covariance. Because the ensemble variance is not directly related to the innovation (observation value minus the ensemble mean value), there is no reason to expect that the pattern of sensitivity values will be the same as that for the observation impact values. Therefore, decisions on where to target with CLAMPS were made subjectively after considering the whole of the observation impact values and the sensitivity values among several variables.

4. Assimilation experiment methods

a. Data assimilation and analysis system

The configuration of the experiments are summarized in Fig. 4 and are described next. Version 3.6.1 of WRF-ARW is used for the data assimilation experiments with the same configuration (Table 2) as the NCAR real-time ensemble that was used to derive the ESA fields, except an ensemble of 80 members is used for the forecasts here rather than the 30 members used in real time. For each case, a 15-km grid with a one-way 3-km nest is initialized from the real-time NCAR-generated 15-km analyses valid at 0600 UTC. Forecasts from the 0.5° GFS 0000 UTC cycle with WRF-VAR perturbations (Torn et al. 2006; Barker et al. 2012) provide lateral boundary conditions for the 15-km grid. The 15-km analysis and the downscaled analysis on the 3-km grid are integrated together such that the 15-km grid provides lateral boundary conditions for the 3-km grid. The 3-km grid is centered near the location of CLAMPS for each case (Fig. 1).

After a 6-h forecast spinup period starting at 0600 UTC, the ensemble adjustment Kalman filter (EAKF; Anderson 2003) encoded in the Data Assimilation Research Testbed (DART) software (Anderson et al. 2009) is used to assimilate observations into the 80 model states starting at 1200 UTC. Observations are assimilated every hour from 1200 to 1700 UTC then
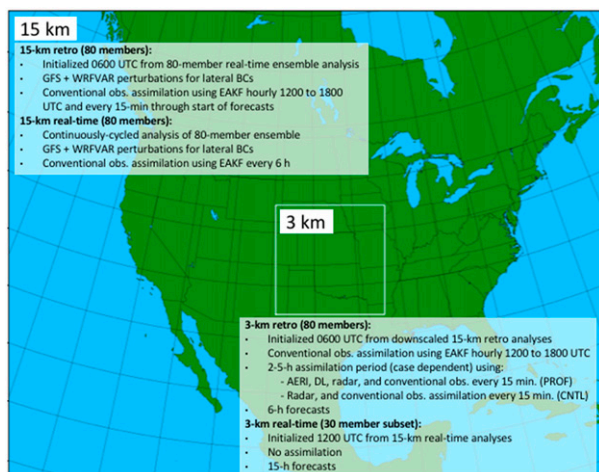
FIG. 4. Summary of the model configurations described in the text. The 15-km and 3-km domain sizes are the same for both the real-time and retrospective experiments. The 3-km domain shown is for the 25 May 2016 retrospective case but was moved as needed for each case.

every 15 min for a case-dependent 2–5-h period starting at 1800 UTC. The analysis from the final assimilation cycle, the time of which varies from 2000 to 2300 UTC among the cases (Table 3), is used to initialize 6-h forecasts produced on the 15-/3-km grids for all 80 members. Observations assimilated in both the 1-h and 15-min assimilation periods include t, t_d, u, v, and surface altimeter from radiosondes and surface stations (METAR, marine, and Oklahoma Mesonet), as well as t, u, and v from the Aircraft Communications Addressing and Reporting System (ACARS). Specified observation errors follow NCEP statistics as in Romine et al. (2016). Following Wheatley et al. (2015), WSR-88D radar reflectivity and radial velocity observations inside the 3-km domain are analyzed every 15 min onto a 6-km grid using a one-pass Barnes's analysis and are only assimilated in the 15-min assimilation period. Radial velocity observations are only assimilated if the gridded reflectivity observation is ≥10 dBZ. Finally, clear-air reflectivity observations are assimilated to suppress spurious storms, but all reflectivity observations <0 dBZ are first set to 0 to mitigate large analysis increments (Wheatley et al. 2015). Specified errors for reflectivity (radial velocity) are 5 dBZ ($2 m s^{-1}$). More details of the assimilation configuration are provided in Table 4.

b. AERI and DL assimilation experiments

AERIoe retrievals of t and t_d (converted from q), and DL retrievals of u and v are only assimilated in the 15-min assimilation period. Because AERIoe retrievals of t and q lose accuracy in cloudy air (defined to have LWP ≥ $5 g m^{-2}$; Turner and Löhnert 2014), only the AERIoe retrievals below cloud base are retained for

assimilation (cloud base height is determined from the zenith-pointing scans of the DL). The t and t_d retrievals are cut off at 2 km AGL because the vertical resolution of the profiles above those heights is often low, particularly for q (Fig. 5a). For the DL retrievals, only those that fit a sine function to the Doppler velocity curve with a root-mean-square difference ≤1 m s^{-1} are retained as a way to eliminate the more questionable VAD estimates. Furthermore, only retrievals that were obtained with sufficient aerosol backscatter, measured by a signal-to-noise ratio ≥ −20 dB (Pearson et al. 2009), are retained. Finally, the few remaining DL retrieval points above 2 km AGL are ignored to match the cutoff height of the AERI retrievals.

After the QC steps above, a temporal Gaussian filter with an e-folding time of 3 min is applied to both the AERIoe and DL retrievals to create lightly smoothed profiles every 15 min to match the time interval of the second assimilation period. After smoothing in time, the DL wind retrievals are smoothed lightly in the vertical (with an e-folding distance of 60 m). Examples of smoothed retrievals are shown in Fig. 6. As noted earlier, AERIoe results in smooth vertical profiles that reflect the rapid loss of independent information content with height (Turner and Löhnert 2014). This is accounted for in the assimilation by using AERIoe retrieval points that are spaced according to the vertical resolution of the profiles estimated in the retrieval (Fig. 5a). For the DL retrievals, the lightly smoothed wind profiles are linearly interpolated to the 40 vertical levels in the model.

Measurement uncertainties are quantified directly by AERIoe as described in section 2a. However, this retrieval error does not include representativeness error resulting from unresolved scales and processes. The standard deviation of representation error for the AERIoe retrievals is specified to be 2 K for both t and t_d and the total observation error (the value used to specify observation error variance in the assimilation) is then the sum of the representation error and the height-dependent measurement error returned by AERIoe. Likewise, the representation

TABLE 2. Primary WRF-ARW options.

Parameter	Outer domain	Nested domain
Horizontal grid	415×325, $\Delta x = 15$ km	381×381, $\Delta x = 3$ km
Vertical grid	40 levels, ptop = 50 hPa	Same
Time step (large)	60 s	12 s
Cumulus scheme	Tiedtke	None
PBL scheme	MYJ	Same
Microphysics	Thompson	Same
LW radiation	RRTMG	Same
SW radiation	RRTMG	Same
Land surface scheme	Noah	Same

TABLE 3. Periods of the AERIoe and DL retrievals used for the PROF experiments resulting in the number of assimilation cycles shown, and a description of the convective mode that later occurred (if any) in the environment that was sampled from the locations shown in Fig. 1.

Date	AERI/DL assimilation period (UTC)	No. of cycles	Convective mode
16 May 2016	1945–2300	14	High-precipitation discrete supercells
23 May 2016	1830–2300	19	Low-precipitation to classic discrete supercells
25 May 2016	1915–2100	8	Short-lived, low-precipitation supercells
26 May 2016	1845–2100	10	Mix of supercells/quasi-linear convective systems
27 May 2016	1845–2000	6	Discrete supercells
22 May 2017	1800–2000	9	Discrete supercells
25 May 2017	1815–2000	8	Discrete supercells
26 May 2017	2015–2200	8	None
27 May 2017	1945–2100	6	Mix of discrete supercells and bowing line segments
31 May 2017	1815–2000	8	Mix of multicells/discrete supercells
11 Jun 2017	2030–2300	11	None
12 Jun 2017	1830–2000	7	Cluster of supercells

error for DL u and v is set to $1.5 \, \mathrm{m \, s^{-1}}$, which is added to the height-dependent measurement error returned by the DL VAD retrieval. Statistics of the total observation error distributions for t, t_d, and u/v are shown in Fig. 5b.

At least 90 min of simultaneous AERI and DL data were collected for each of the 12 cases, allowing for at least 6 cycles in which the processed AERIoe and DL retrievals are assimilated (Table 3). The control (CNTL) experiments assimilate all observations except for the AERIoe and DL retrievals. The experiments that add the AERIoe and DL retrievals are referred to as the PROF experiments. The time when the AERIoe and DL retrievals are first assimilated is determined by the start of the sampling period, and the time of the final assimilation cycle for each case is chosen so that a 2-h period of short-term forecasts (covering 1–3 or 2–4 h) can be made to match the valid time of the forecast variables evaluated in ESA. For example, MVKE in the ESA fields for the 25 May 2017 case is evaluated over a valid time of 2100–2300 UTC (9–11-h forecasts from the real-time NCAR ensemble; Fig. 2b). Therefore, the final assimilation cycle for the 25 May 2017 experiments is

chosen to be 2000 UTC to produce 1–3-h forecasts valid at 2100–2300 UTC. This design provides a way to evaluate the efficacy of the real-time ESA guidance, as well as the ability of the AERI and DL to improve forecasts of convective initiation and early convective evolution on time scales applicable to short-term, convective forecasting applications (e.g., the Warn-on-Forecast initiative; Stensrud et al. 2009; Lawson et al. 2018).

c. Verification

As in many past evaluations of convection-allowing models (e.g., Romine et al. 2013; Coniglio et al. 2016), forecasts of convection are verified using the fractions skill score (FSS) and fractions Brier score (FBS) (Roberts and Lean 2008; Schwartz et al. 2010) given by

$$\mathrm{FSS} = 1 - \frac{\mathrm{FBS}}{\mathrm{FBS_{worst}}}, \qquad (2)$$

where

$$\mathrm{FBS} = \frac{1}{N} \sum_{n=1}^{N} [\mathrm{NP}_{f(i)} - \mathrm{NP}_{o(i)}]^2, \qquad (3)$$

TABLE 4. Primary DART options. Choices for the localizations follow Wheatley et al. (2015) (radar), Coniglio et al. (2016) (radiosonde), and Sobash and Stensrud (2013) (mesonet). Horizontal localization for the AERI, DL, and ACARS observations were made smaller than that used for the radiosonde observations because they are assimilated more frequently.

| Parameter	Value
Filter type	Ensemble adjustment Kalman filter
Adaptive inflation	1.0, 0.6 (initial mean, spread)
Localization type	Gaspari–Cohn
Outlier threshold	3.0
Radiosonde horizontal (vertical) localization half-width	240 (4) km
ACARS horizontal (vertical) localization half-width	120 (4) km
AERI and DL horizontal (vertical) localization half-width	120 (4) km
METAR/maritime horizontal (vertical) localization half-width	60 (4) km
Surface mesonet horizontal (vertical) localization half-width	30 (4) km
Radar horizontal (vertical) localization half-width	9 (3) km
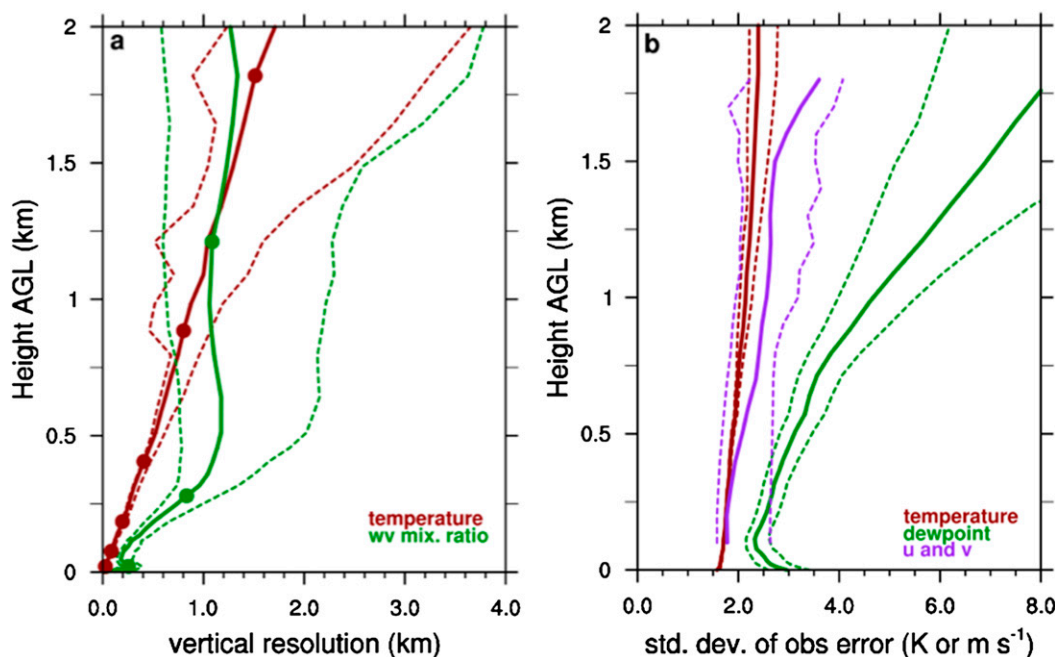
FIG. 5. Vertical profiles of (a) the vertical resolution (km) of the AERI t (red) and q (green) retrievals and (b) the total observation error (standard deviation) of t, t_d, u, and v (purple) specified in the assimilation. Solid (dashed) lines depict medians (10th and 90th percentiles) across all 12 cases. Only points with at least 30 samples that meet the QC criteria described in the text are included. Filled circles on the median curves in (a) depict the heights of a typical set of AERIoe retrieval points that are assimilated as determined by the thinning procedure described in the text.

and

$$\text{FBS}_{\text{worst}} = \frac{1}{N}\left[\sum_{n=1}^{N}\text{NP}_{f(i)}^2 + \sum_{n=1}^{N}\text{NP}_{o(i)}^2\right]. \qquad (4)$$

The neighborhood probability (NP) is the fraction of grid points within a surrounding neighborhood that contains values exceeding a threshold of some field (to allow for some mismatch between observations and forecasts to still be counted as a ''hit''). The subscripts $f(i)$ and $o(i)$ represent the ith grid box in the forecast and observed fields, respectively, and N is the number of grid points in the verification domain. FSS values range from 0 (no skill) to 1 (perfect),[3] whereas an FBS of 0 is considered perfect and increasing FBS values indicate a decreasing correspondence between the forecasts and observations. FBS$_{\text{worst}}$ is achieved when there is no overlap of nonzero fractions and represents a low-accuracy reference forecast that is needed to assess skill (through the FSS).

The forecast field used in the verification is simulated composite reflectivity and the observed field is the NSSL Multi-Radar Multi-Sensor (MRMS) 0.01° by 0.01°

analysis of composite reflectivity (Zhang et al. 2016) interpolated to the 3-km WRF-ARW grid. Using a threshold[4] of 35 dBZ, the NPs for each member are computed from these reflectivity fields and are then averaged across all ensemble members, which is then used to compute the FSS as in Schwartz et al. (2010). The verification domain covers only the area of convection that evolved within and near the ESA region to prevent FSS sensitivities that can occur when the ''wet area'' ratio of the domain is small (Mittermaier and Roberts 2010) and to mitigate the influence of errors from convection in the 3-km domain that does not evolve within the environment that was sampled.

Although many reflectivity thresholds and neighborhood sizes were tested, we focus on results using 35 dBZ and a square neighborhood of 8 grid cells giving a neighborhood size of ~25 km by 25 km as this combination is found to provide a good representation of skill that matches visual inspection of the differences. The interpretation of the relative differences in FSS is not

---

[3] FSS values ≥~0.5 may represent forecasts with ''useful'' skill (Roberts and Lean 2008), but the absolute values of FSS are less important here than changes in FSS between experiments.

[4] To examine the influence of bias on the results, tests were performed using the percentile method to define the reflectivity thresholds (Mittermaier and Roberts 2010). The FSS and FBS comparisons were found to be nearly identical when using this bias-adjustment technique versus using a constant 35-dBZ threshold.
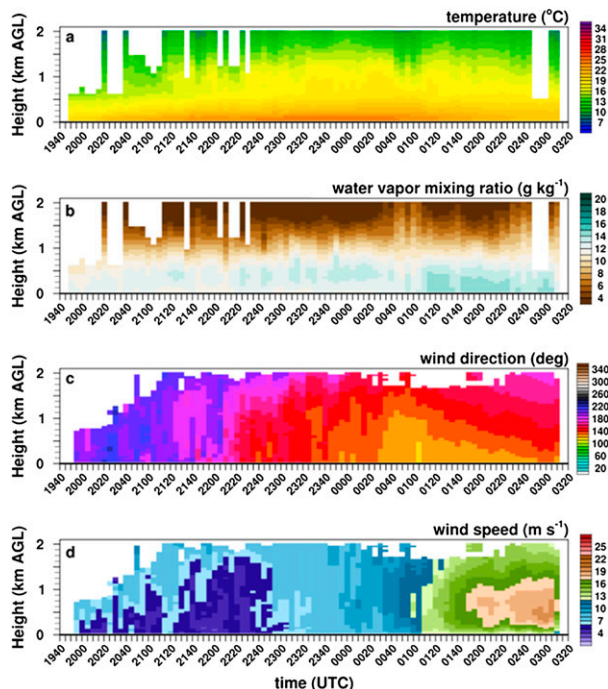
FIG. 6. Time–height cross sections of postprocessed AERIoe retrievals of (a) temperature (b) water vapor mixing ratio and DL retrievals of (c) wind direction and (d) wind speed made from 1952 UTC 16 May to 0314 UTC 17 May 2016 from the location indicated on Fig. 1. The profilers are effective at sampling sub-hourly changes in the boundary layer state, like the increase in moisture below 500 m AGL beginning at 0110 UTC seen in (b) and the substantial backing and strengthening of the winds after 0030 UTC seen in (c) and (d).

sensitive to these choices. This neighborhood reduces the influence of errors with spatial scales close to and smaller than the smallest resolvable scale of the grid ($\leq \sim 4$–6Δ; Skamarock 2004) while retaining scales immediately larger than individual thunderstorms. Inclusion of these relatively small scales in the verification is important for the goal of assessing the impact of observations on ensemble forecast systems that aim to provide skillful storm-scale forecast guidance (e.g., Stensrud et al. 2009; Wheatley et al. 2015; Lawson et al. 2018).

5. Findings

a. 25 May 2017 case

Assimilation of the AERI and DL retrievals results in noticeable changes to the ensemble mean fields in the CBL for all cases. The typical changes to the ensemble mean in the CBL, regarding both the spatial extent of the changes (governed by the choice of covariance localization; Table 4) and the magnitude of the changes, are illustrated in Fig. 7 for the 25 May 2017 case. The flow-dependent and nonisotropic nature of the EAKF

covariances results in a pattern of cooling and moistening in the CBL that reflects the shape of a dryline to the west and a cold front to the north (Figs. 7e,f). The pattern reflects a westward shift of the dryline in the PROF ensemble mean, which is also reflected in the easterly vector wind difference. This westward shift results in a dryline position closer to where it was observed. This westward shift to the dryline, and the associated convergence in the CBL, results in more divergence near the CLAMPS location in the PROF experiment with a slight cooling and drying also noted there. The result is an environment more supportive for convection farther west, within far-western Kansas, and an environment less supportive of convection near CLAMPS.

Differences in neighborhood ensemble probabilities (NEPs)[5] for the 25 May 2017 case are small through the first 2-h of the forecasts (Figs. 8a–c). After 2 h, as the convection interacts with the modified environment, NEPs that are 5%–10% higher in the PROF experiment emerge near the location of the supercell over northwestern Kansas and NEPs that are 5%–15% smaller in the PROF experiment emerge to the east of this supercell where no storms were observed (Figs. 8e,f). This westward shift of NEPs in the PROF experiment to a location over the supercell is consistent with the westward shift in the dryline closer to where it was observed. This dipole in NEP differences continues through 4 h, with increases of NEPs up to 10% that overlap with the eastern portion of what evolved into a small cluster of supercells. Furthermore, decreases in NEP continue to be seen up to 15% in the PROF experiment where no storms were observed to the northeast of the supercell cluster (Fig. 8i). By 5 h, the pattern of the differences becomes much less coherent and the magnitudes of the NEP differences become small again (Fig. 8l).

Consistent with the negligible NEP differences through 2-h forecasts, the FSS is nearly identical for both experiments through about 2 h (Fig. 9). Also consistent with the NEP differences shown in Fig. 8 for 3- and 4-h forecasts, the FSS is larger for the PROF experiment at these times, with up to 95% confidence in the differences at 3 h (Fig. 9). The FSS differences become small again after 4 h (Fig. 9).

b. Impacts over all twelve cases

The NEP differences provide a helpful way to visualize the impacts on the forecasts and are presented for all 12 cases (Fig. 10) at a time when the FSS differences (PROF minus CNTL) were large for that case. While the NEP

---

[5] NEPs are simply the ensemble mean of the neighborhood probabilities as described in Schwartz et al. (2010).
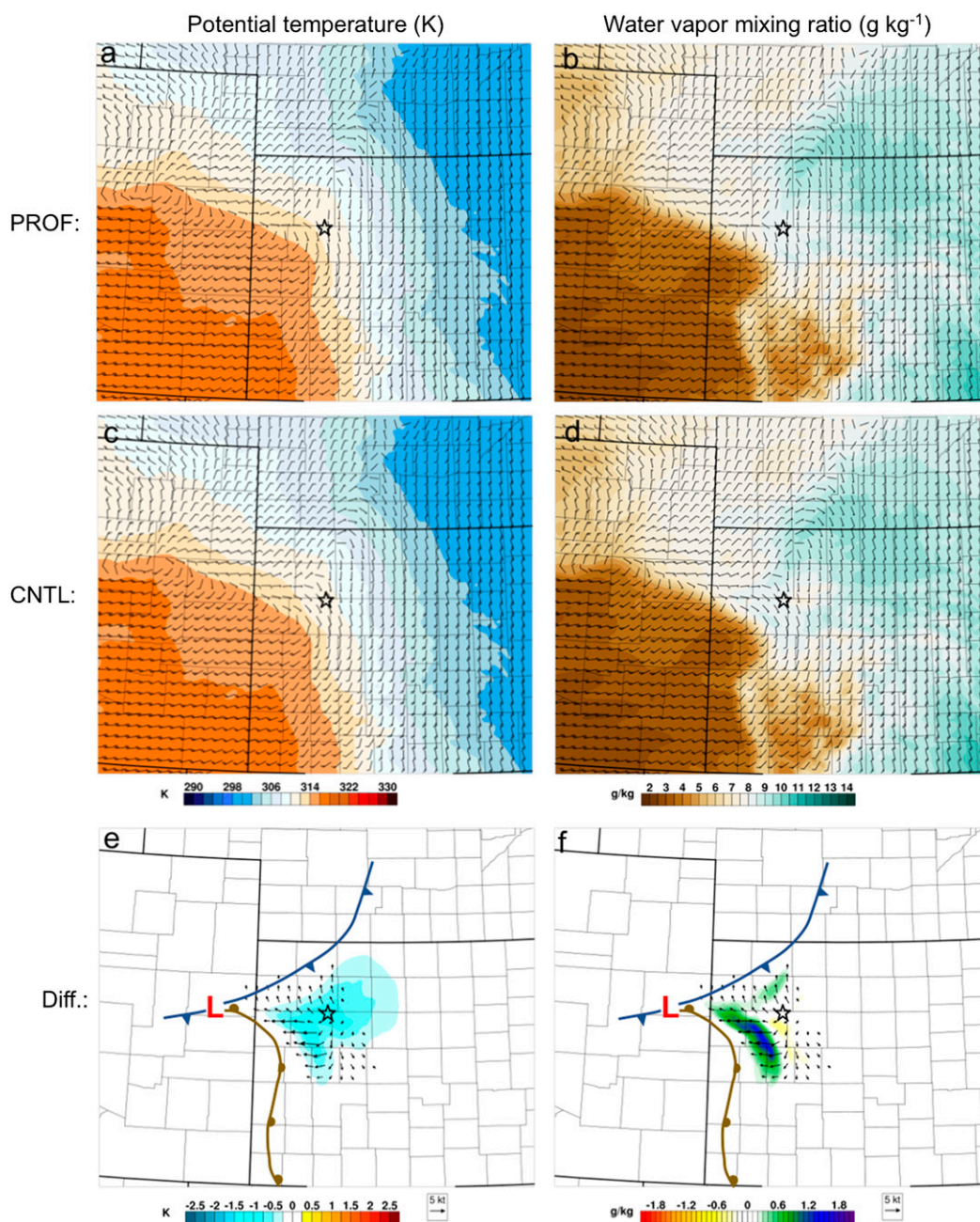
FIG. 7. (a)–(f) Ensemble mean potential temperature (K), water vapor mixing ratio (g kg^{-1}), winds (full barbs are 10 kt, half barbs are 5 kt in a-d), and their difference (PROF − CNTL) for the final assimilation cycle valid 2000 UTC 25 May 2017 on model level 6, or approximately 650 m AGL. Symbols in (e) and (f) represent the manual surface analysis of MADIS METAR and mesonet observations and are drawn independent of the ensemble analyses.

difference magnitudes are rather small—mostly <20%—the positive differences (higher probabilities in the PROF experiments) tend to overlap or are close to observed storms and the negative differences (lower probabilities in the PROF experiments) tend to be in areas devoid of storms.

Figure 11 presents the FSS for all twelve cases. Presenting the FSS differences for each case more
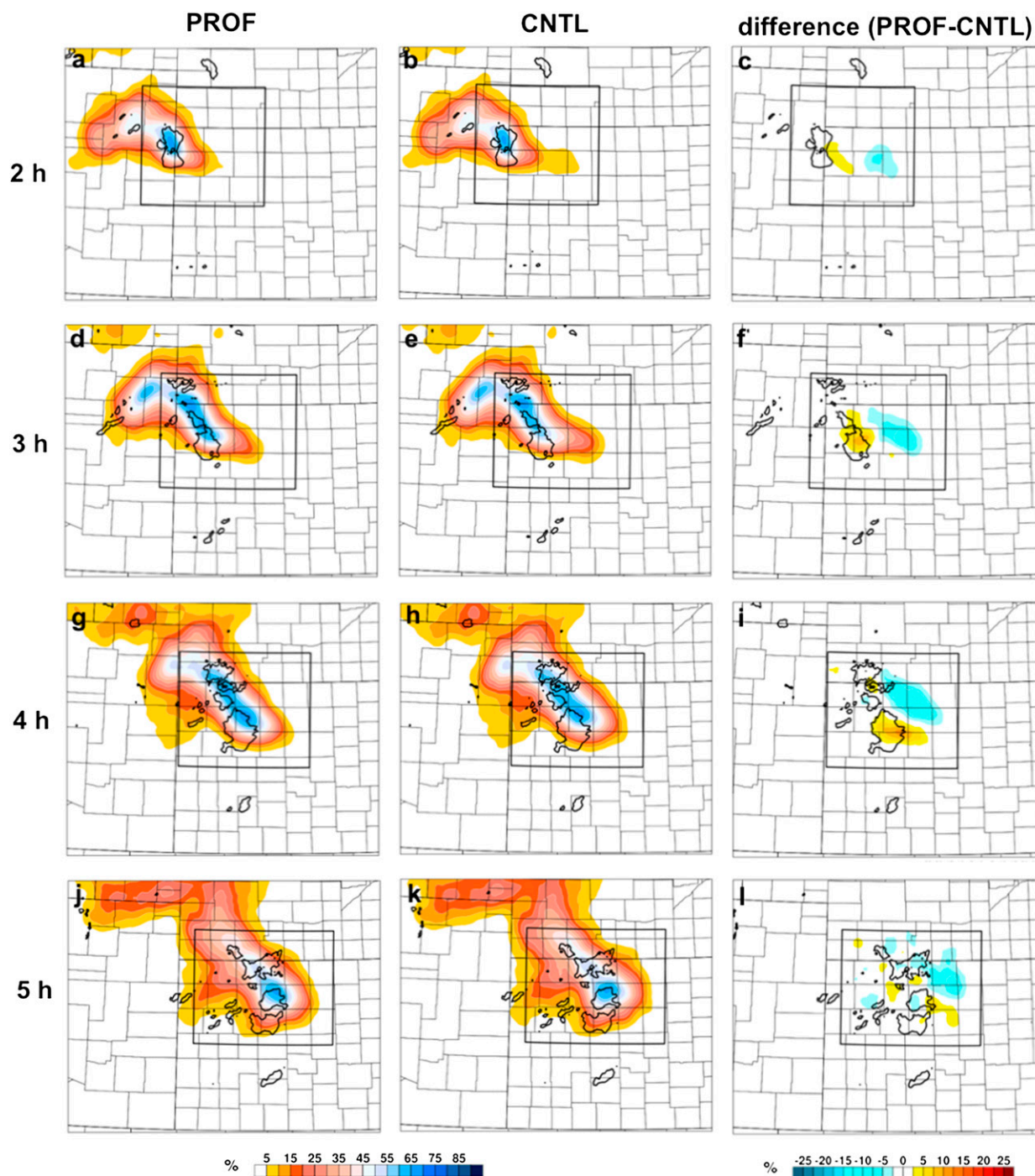
FIG. 8. Neighborhood ensemble probabilities of simulated composite reflectivity ≥35 dBZ (color shading) using ~25 km by 25 km neighborhoods for the (left) PROF experiment, (middle) CNTL, and (right) their difference for the 25 May 2017 case for 2-, 3-, 4-, and 5-h forecasts (valid 2200 UTC 25 May to 0200 UTC 26 May 2017). The black contours are the MRMS analysis of composite reflectivity ≥35 dBZ. The black rectangle encloses the area used to compute the FSS.

effectively conveys the overall impacts of assimilating the AERI and DL retrievals compared to a single aggregate across all cases because the forecast times when the impacts are seen vary across the cases (Fig. 11)

resulting in muted impacts when aggregated. The red (blue) shading in Fig. 11 displays when the PROF minus CNTL (CNTL minus PROF) FSS is ≥0.01. The 95% confidence intervals can be used to assess the confidence
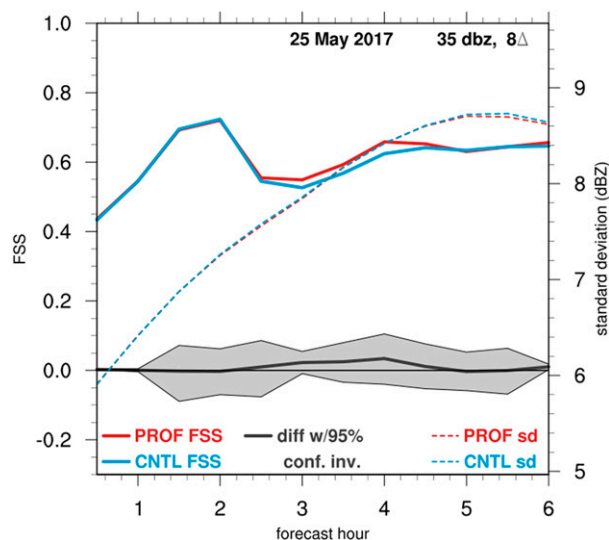
FIG. 9. The fractions skill score (left axis) for the PROF experiment (solid red), the CNTL experiment (solid blue), and their difference (solid black) computed in the moving area indicated in Fig. 8 for a neighborhood size of 8 by 8 grid points (~25 km by 25 km). The 95% confidence intervals shown in the shaded area around the FSS difference is computed using a bootstrap method with 1000 resamples outlined in Hamill (1999) where the ensemble members are used as the independent samples in both sets. Dashed lines indicate the ensemble standard deviation in simulated composite reflectivity (right axis) for both experiments.

in those differences; the closer the edge of the interval is to zero, the more confidence there is that the differences are statistically meaningful. Although most of the differences have confidence intervals that overlap zero, Fig. 11 shows that there are many more times when the FSS is larger for the PROF experiments than for the CNTL experiments. This is evidence that the AERI and DL assimilation is providing frequent positive impacts, although the time period, duration, and magnitude of the positive impacts do vary substantially across the cases. A time when positive forecast impacts appears to be somewhat consistent is around 3 h into the forecasts. The two cases that do not show improvement around 3 h are the 27 May 2016 and 26 May 2017 cases; the background environment was changed very little by the AERI and DL assimilation in the former (because of an already accurate background analysis and only six assimilation cycles) and no convection was observed in the area in the latter.

Some mention of the near-zero FSS for the 26 May 2017 case (Fig. 11h) is warranted. The zero FSS reflects times when there were no storms observed in the evaluation domain for this case. When there are no observed storms, the FSS will be 0 for both forecasts regardless of the forecast NPs for either ensemble because $NP_{o(i)}$ is 0, resulting in a FSS of 0 [as can be seen in Eqs. (2)–(4)]. However, in the 26 May 2017 case the PROF experiment

produces more spurious storms than the CNTL experiment, and thus should score worse than the CNTL experiment. To better compare the forecasts in cases with few or no storms in the evaluation domain, Fig. 12 presents the FBS for four such cases since higher nonzero NPs will result in higher FBS values (a poorer score) regardless of the occurrence of observed storms [see Eq. (3)]. The FBS is indeed lower for the CNTL experiment for the 26 May 2017 case (Fig. 12c). For the other cases, little change in the interpretation of the relative accuracy of the two experiments is seen when using the FBS, as the PROF experiment also scores higher from 1 to 6 h for the 22 May 2017 and 11 June 2017 cases (two cases in which the PROF experiment reduced the number of spurious storms in the area), and the differences in the PROF and CNTL experiments remain relatively small for the 25 May 2016 case (although the positive differences are seen at different times).

c. Relationship of impacts to ESA fields

Overall, no robust relationships between the real-time ESA fields to actual forecast impacts are found; the magnitude of the observation impact values did not relate linearly to the magnitude of the changes in either forecast skill or forecast variance (the standard deviation of the forecasts can be seen in Fig. 11). Likewise, no clear statistical relationships between sensitivity to q, u, and v averaged in the lowest 1 km and the ensemble mean reflectivity are found. This is not surprising because multiple observation types from the profiler retrievals were assimilated (t, t_d, u, and v) and the sign of the sensitivity of ensemble mean reflectivity to each individual field may not be the same; that is, assimilation of t observations may pull the ensemble mean one way and assimilation of u observations might pull it the other way. However, there is some evidence that the sensitivity for potential temperature averaged in the lowest 1 km AGL ($\overline{\theta}_{1km}$) to the ensemble mean reflectivity (Table 5) may have a larger influence than the other variables. In general, Eq. (1) predicts the sign and magnitude of the change in the forecast variable J to a change in a state variable x, so for a positive sensitivity ($\partial J/\partial x > 0$), a positive (negative) change in x should result in a positive (negative) change in J. Likewise, for a negative sensitivity ($\partial J/\partial x < 0$), a negative (positive) change in x should result in a positive (negative) change in J. As seen in Table 5, the $\overline{\theta}_{1km}$ sensitivity predicted the sign of the actual difference in ensemble mean reflectivity correctly in 11 out of 12 cases (although it is probably chance that the correct sign was predicted for the 22 May 2017 case given the very small change in $\overline{\theta}_{1km}$ of -0.02 K). For example, the positive $\overline{\theta}_{1km}$ sensitivities near CLAMPS for the 25 May 2017 case (Fig. 3a), combined with the actual decrease in $\overline{\theta}_{1km}$ resulting from
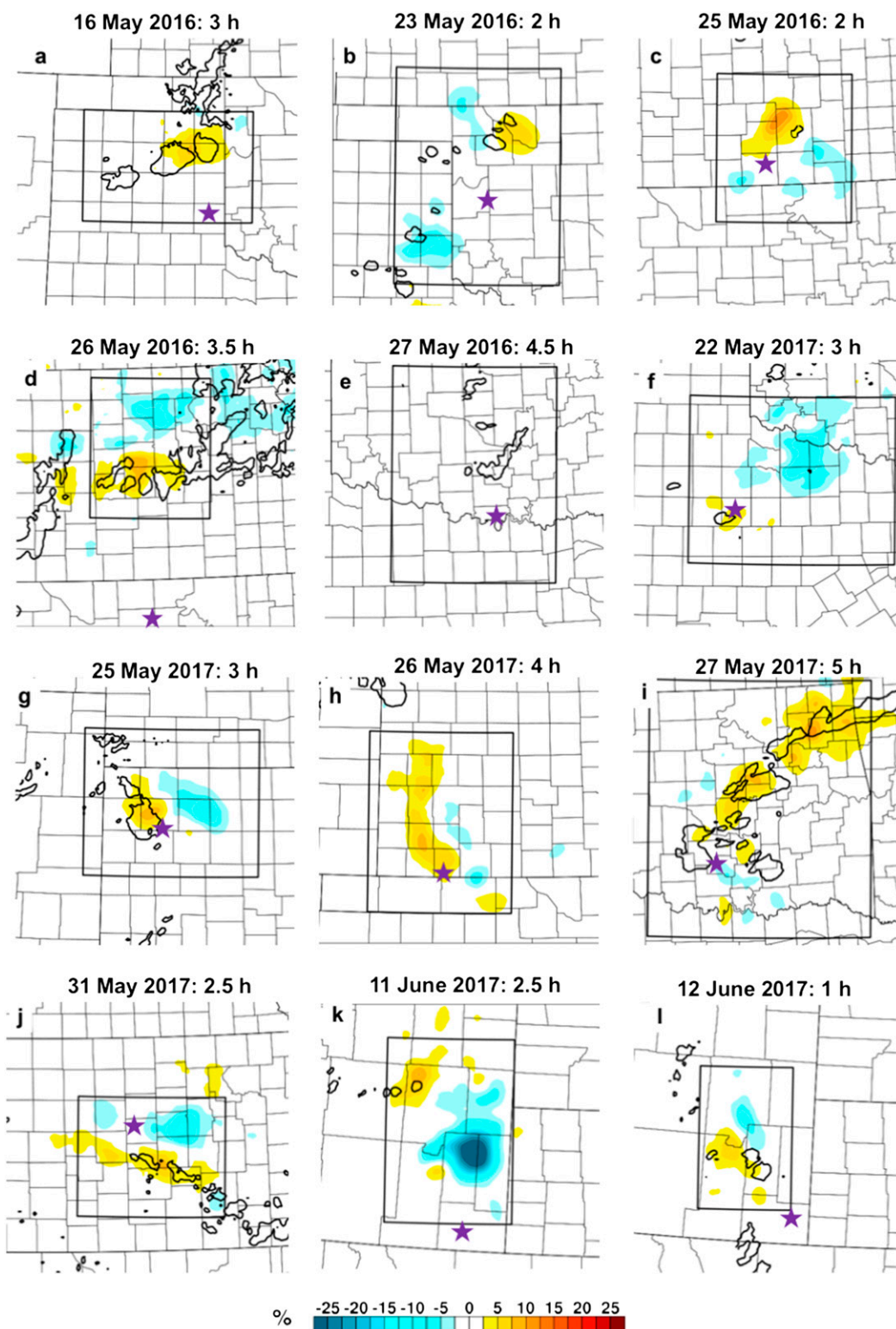
FIG. 10. The difference in neighborhood ensemble probabilities (NEP) between the PROF experiment and the CNTL experiment for ~25 km by 25 km neighborhoods for all 12 cases for a forecast time with the largest positive difference in FSS (shown in the title of each panel). The black contours are the MRMS analysis of composite reflectivity ≥35 dBZ. All difference magnitudes for the 27 May 2016 case shown in (e) are less than 2.5% but it is included for completeness. Improvements in the forecasts can be visualized by positive differences (warm colors) that overlap or are close to the observed storms and/or negative differences (cool colors) that are not close to observed storms. The rectangle encloses the area used for the FSS computations and the purple star indicates the location of CLAMPS earlier in the day.
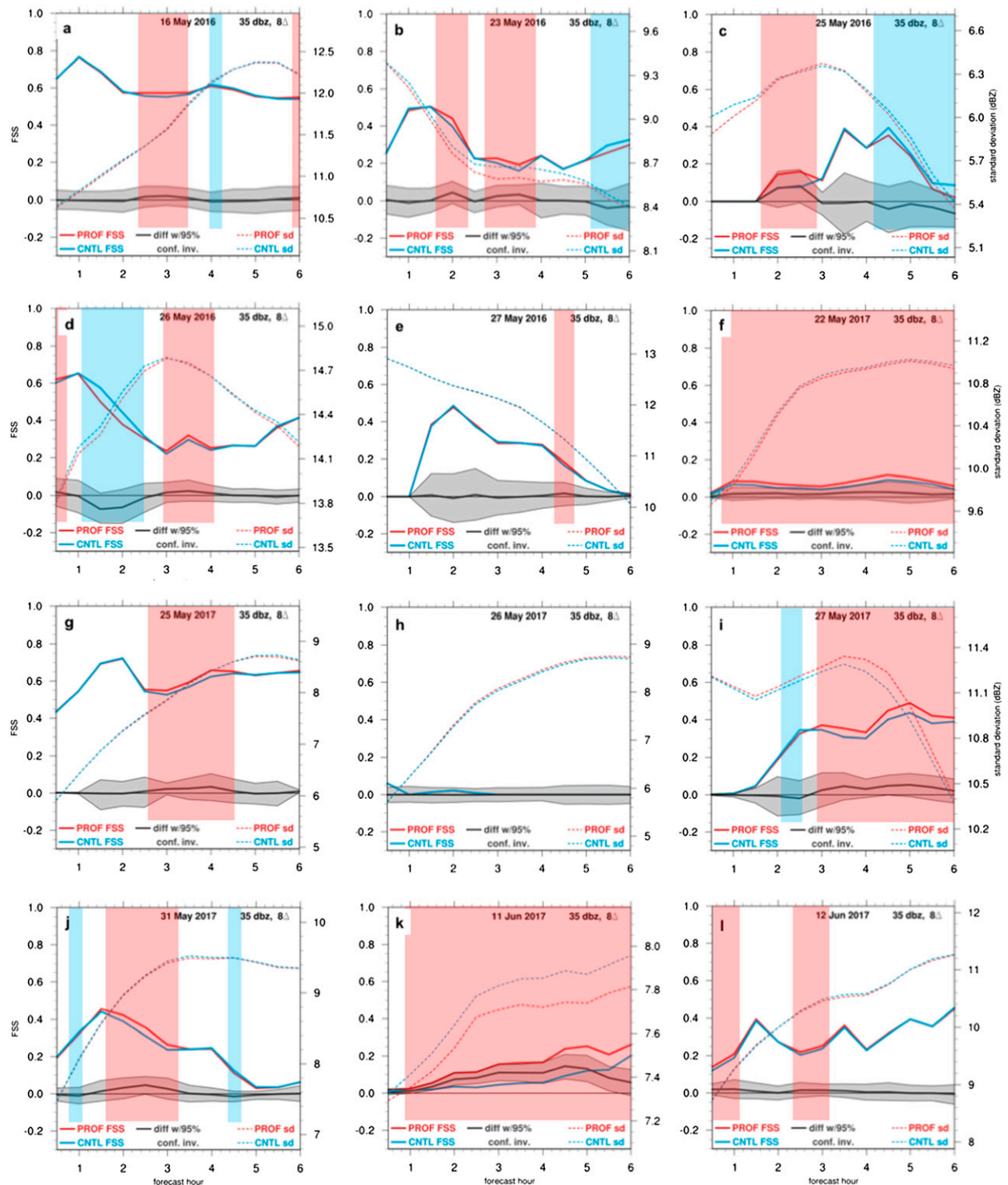
FIG. 11. As in Fig. 9, but for all twelve cases. Red (blue) shading indicates the times when the FSS for the PROF (CNTL) experiment is at least 0.01 larger than the FSS for the CNTL (PROF) experiment, thereby indicating higher skill for the PROF (CNTL) experiment.
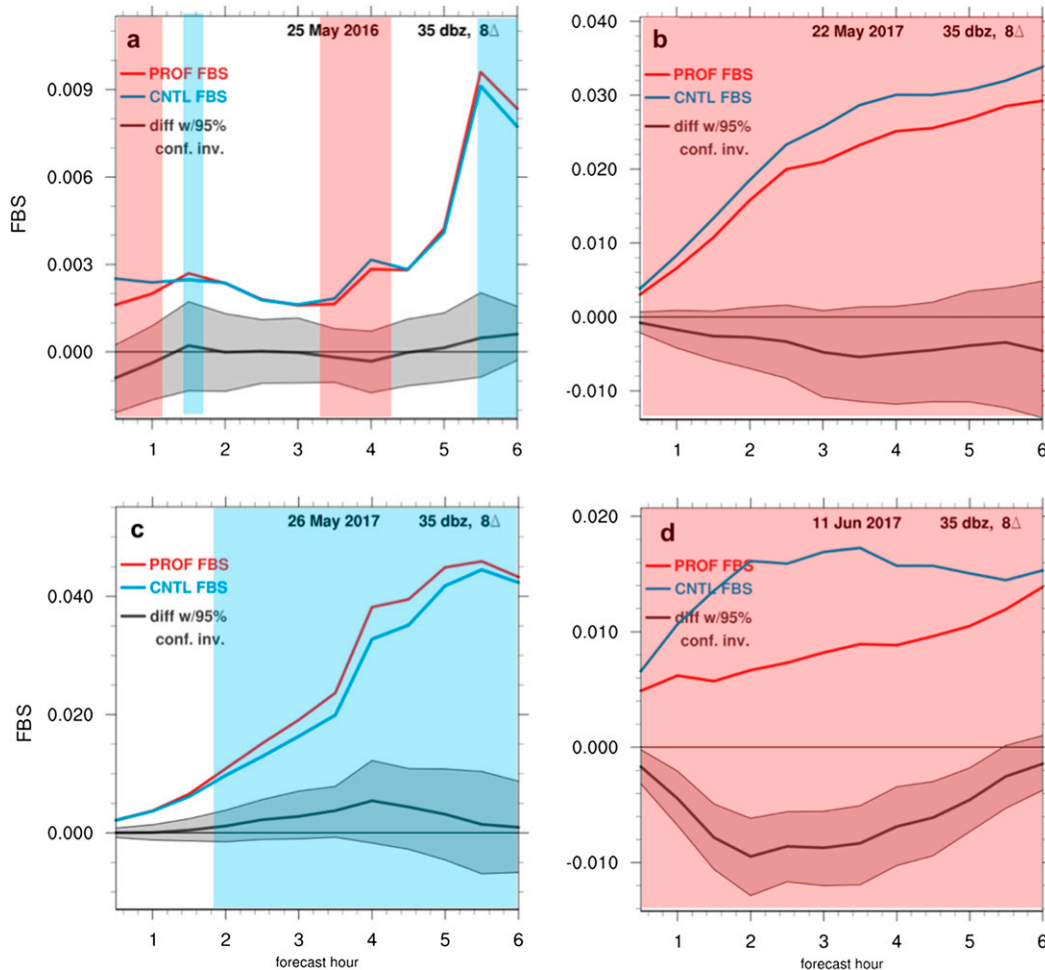
FIG. 12. As in Fig. 9, but for the FBS for the four cases with few or no observed storms in the evaluation domain. For the FBS, values closer to zero indicate better scores, so red (blue) shading indicates the times when the FBS for the PROF (CNTL) experiment is at least 0.01 smaller than the FBS for the CNTL (PROF) experiment, thereby indicating higher accuracy for the PROF (CNTL) experiment.

the assimilation of the AERI and DL retrievals (Fig. 7e and Table 5), predicts that the ensemble mean of MVKE will be lower in the area shown in Fig. 2a in 1–3-h forecasts. Using the simulated reflectivity as a proxy for MVKE, we see that indeed the ensemble mean reflectivity was lowered in this approximate area (Table 5), which can be seen by the larger magnitudes of the negative NEPs than those for the positive NEPs in Figs. 8c and 8f. An indication that these results may not be robust is that the one case (12 June 2017) in which the $\overline{\theta}_{1km}$ sensitivity did not predict the change in ensemble mean reflectivity correctly had by far the largest $\overline{\theta}_{1km}$ sensitivity values. Only the sensitivity for lowest-1-km-averaged u (out of t, t_d, u, and v) predicted the sign of the ensemble mean reflectivity correctly for this case.

The interesting results for potential temperature sensitivity aside, there are many possible reasons for the

lack of identified relationships between the ESA fields and the forecast impacts. Attempts were made to locate CLAMPS in an area with at least moderate sensitivity or observation impact as shown in Fig. 3, but this was not always possible because of the numerous logistical challenges of mobile ground-based observation targeting. Furthermore, the forecast fields used for the sensitivity calculations were averaged in space and time prior to the development of deep convection in an attempt to produce more robust and effective sensitivity fields (Torn and Romine 2015). However, the averaged sensitivity fields still contained large mesoscale gradients and noise, and likely still contained sampling error related to a relatively small ensemble (30 members) (e.g., Fig. 3). Therefore it was sometimes difficult to relate the sensitivity fields to physical processes like that shown for the 25 May 2017 case in section 3. Of course sampling

TABLE 5. (left) The near-CLAMPS 0–1-km AGL mean potential temperature ($\overline{\theta}_{1km}$) sensitivity to the 1–3-h forecast ensemble mean simulated reflectivity averaged in the ESA metric region. (middle) The actual near-CLAMPS ensemble mean difference in $\overline{\theta}_{1km}$ (PROF-CNTL) resulting from the assimilation of the AERI and DL observations. (right) 1–3-h forecast ensemble mean simulated reflectivity difference (dBZ) averaged in the ESA metric region. The sensitivity values have units of K^{-1} since the base sensitivity values of dBZ K^{-1} (in 2016) and m^2 s^{-2} K^{-1} (in 2017) are normalized by the ensemble standard deviation of the forecast metric within the metric area.

Date	6–9-h mean $\overline{\theta}_{1km}$ sensitivity	Actual $\overline{\theta}_{1km}$ difference (K)	Ensemble mean reflectivity difference (dBZ)
16 May 2016	+0.20	+0.56	+0.38
23 May 2016	−0.28	+1.20	−0.34
25 May 2016	+0.04	+0.35	+0.02
26 May 2016	−0.07	+1.03	−1.23
27 May 2016	−0.19	+0.21	−0.07
22 May 2017	+1.08	−0.02	−0.39
25 May 2017	+3.39	−0.61	−0.14
26 May 2017	−2.22	−0.35	+0.34
27 May 2017	+2.20	0.32	+0.19
31 May 2017	+1.16	−0.61	−0.06
11 Jun 2017	+4.87	−1.04	−1.61
12 Jun 2017	+7.40	−0.63	+0.19

error from the limited sample size of cases is another problem. Yet another potential problem is that the experiments—and real-time ESA guidance—were designed to capture the initiation and early evolution of convection in 1–3-h forecasts. Predicting the first storms, an inherently small-scale and often nonlinear process, has long been recognized as the biggest challenge for storm-scale NWP (Lilly 1990; Stensrud et al. 2009). It is not surprising that the skill of convection-allowing models of predicting the timing of convective initiation within tens of minutes and tens of kilometers is still highly limited (Bytheway and Kummerow 2015; Hill et al. 2016; Burlingame et al. 2017). Applying ESA fields to this prediction problem is precarious because they rely on linear relationships. More robust relationships could be hindered further because the ESA fields are generated from 6–9-h forecasts, and small errors in 6–9-h forecasts of the CBL could lead to large errors in forecast metrics related to convection, even those metrics that are averaged over space and time. The coherency in time of the ESA fields was not examined here and should be examined in future studies to assess the general utility of ESA techniques for convective-scale forecasts.

6. Experiment sensitivities

The sensitivity of the results shown in Figs. 10 and 11 was tested in numerous ways, including by varying the

reflectivity threshold, neighborhood sizes, and evaluation domain sizes. Parameters in the EAKF procedure also were varied, including the covariance localization and representativeness errors in the AERI and DL retrievals. There are no noteworthy differences in the interpretation of the results by changing these parameters of the assimilation and verification within reasonable bounds. Regarding the lack of sensitivity to neighborhood size in the verification, this is different than what was found in Coniglio et al. (2016) for forecast impacts to the assimilation of radiosonde observations. In Coniglio et al. (2016), larger differences in FSS were seen among all eight of their cases as the neighborhood size increased from ~25 km by 25 km to ~100 km by 100 km. Here, the FSS differences are similar for both neighborhood sizes. This implies that the improvements to the forecasts in this study tend to be made on smaller spatial scales than those made in Coniglio et al. (2016), which can be seen in Fig. 10 with positive FSS differences remaining rather close to the observed storms if not directly overlapping them. The reasons for these differences in results between the studies is not clear, but could indicate the benefit of having subhourly CBL observations versus observations taken at intervals equal to or greater than hourly as done with the radiosondes in Coniglio et al. (2016).

However, the positive impacts within ~25 km by 25 km neighborhoods seen in this study are often smaller than those seen in Coniglio et al. (2016). More specifically, there were four cases in Coniglio et al. (2016) in which the 95% confidence interval for the ~25 km by 25 km neighborhood FSS differences did not overlap with zero for at least a 1-h period [see Fig. 17 in Coniglio et al. (2016)], whereas there is only one case in this study in which that occurs (Fig. 11k). There are many possible reasons for this difference, one being the relative impact of remotely sensed AERI and DL retrievals versus in situ radiosonde observations, including data resolution, quality, and availability of observations above 2 km. Another possible reason is that in all cases in Coniglio et al. (2016) there were multiple radiosonde observations obtained over a mesoscale network, whereas in this study observations from only one location were made. The Southern Great Plains (SGP) site of the Department of Energy Atmospheric Radiation Measurement (ARM) program now contains a network of five AERIs and DLs spread over north-central Oklahoma (Wulfmeyer et al. 2018). The impacts of retrievals from these systems (and from the temporary network of research AERIs and DLs used for PECAN; Geerts et al. 2017) on forecasts of convection are currently being investigated at NSSL and other institutions to help determine if the skill improvements, and the confidence in the skill improvements,
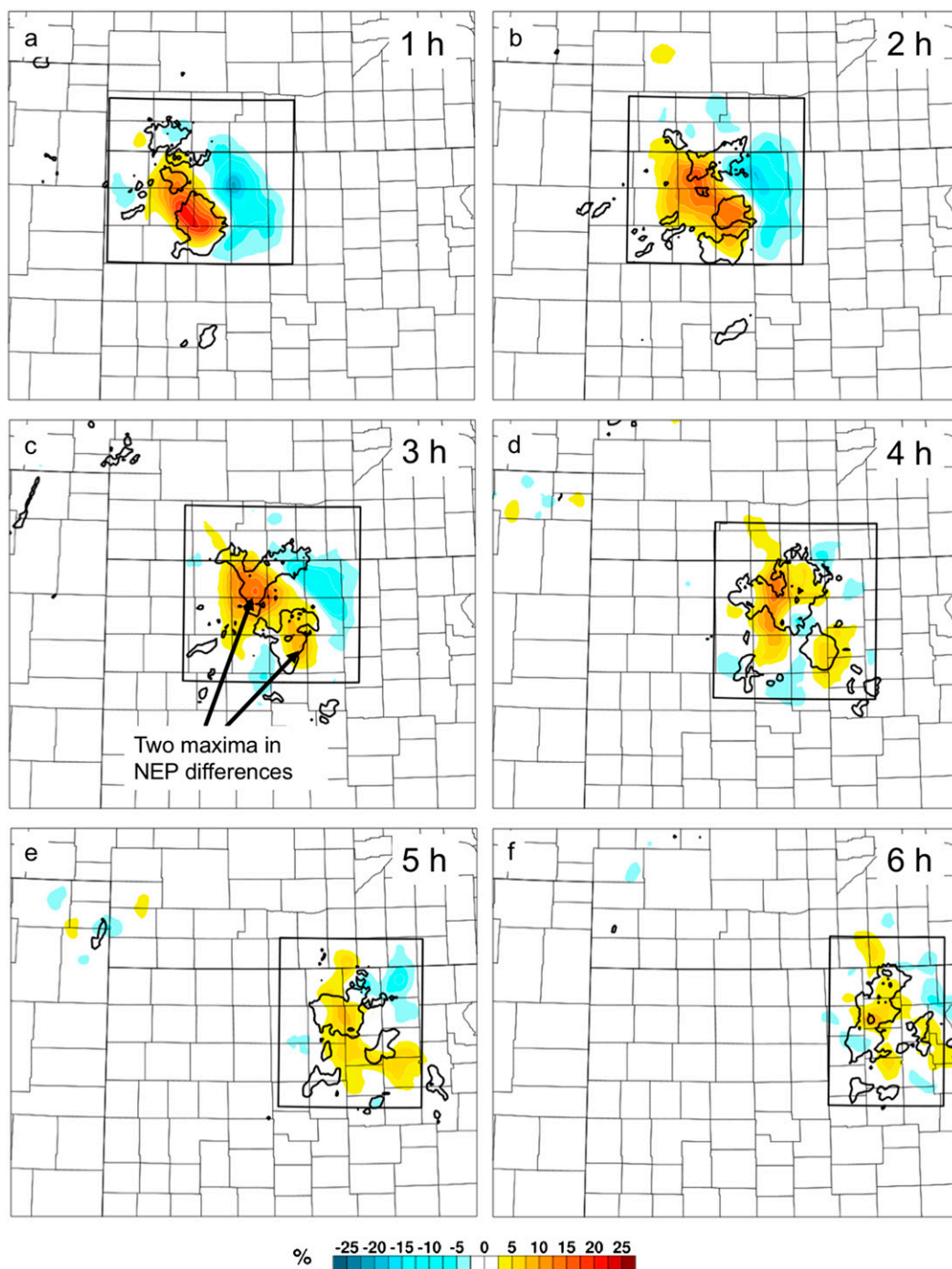
FIG. 13. As in Fig. 8c, but for the 25 May 2017 experiment initialized at 2300 UTC for forecasts every hour from 1–6 h (valid at 0000–0500 UTC 26 May 2017). Two maxima in NEP differences described in the text are pointed out in (c).

would increase if retrievals from a network of profilers are assimilated versus from one targeted system of profilers.

One notable sensitivity of the forecast impacts to the assimilation procedure was found and relates to the number of cycles in which AERI and DL data are assimilated. While this study was designed to focus on convective initiation and early convective evolution, experiments were produced that extended the assimilation period into the afternoon/early evening during the mature stage of convection. For example, an experiment

was performed for the 25 May 2017 case in which data were assimilated through 2300 UTC rather than 2000 UTC giving 20 cycles instead of 8 (Table 3), with storms occurring within 15 of those cycles instead of 4 in the 2000 UTC-initialized case. This longer period of assimilation does little to reduce initial condition errors of the background environment, which reflects the well-calibrated and stable nature of the NCAR ensemble (Fig. 14). However, extending the assimilation period allows more time for storms to evolve and mature within the cycled analysis and contributes to reduced errors for reflectivity and improvements to the spread/error relationship[6] for both reflectivity and radial velocity (Fig. 15).

Despite the better analysis of ongoing storms, the AERI and DL retrievals are still able to make positive impacts to the forecasts in the 25 May 2017 case, and the impacts are even amplified. The additional cycles of AERI and DL retrievals allows for the impacts to become well established in the analysis in the presence of the storms (several other cases not shown also display this behavior). For example, the final analysis at 2300 UTC for the 25 May 2017 case shows more mesoscale structure to the amplified difference fields (Fig. 16). As for the 2000 UTC-initialized case, the differences in NEP continue to show an improvement in the placement of the cluster of supercells (Fig. 8), but the maximum differences are amplified (now 20%–25%) both within the area where NEP is larger for the PROF experiment near the supercell cluster and where the NEP is smaller for the PROF experiment for the spurious storms to the northeast of the supercell cluster (Fig. 13a). This is for the 1-h forecasts, but the later forecasts show that these larger differences in NEP are not just because of the shorter lead time; the differences in FSS for the 3-h forecasts from the 2300 UTC-initialized case are larger than those for the 3-h forecasts from the 2000 UTC-initialized case (cf. Fig. 13c and Fig. 8c). The PROF experiment for the 2300 UTC-initialized case even shows a pattern of increases in NEPs, most well defined at 3 h (Fig. 13c), that reflect the two separate supercell clusters that were observed.

The larger NEP differences produce larger FSS differences for the 2300 UTC-initialized case compared to the 2000 UTC-initialized case (cf. Fig. 9 and Fig. 17). These improvements are made despite an increase in the CNTL FSS for the 2300 UTC-initialized case compared to the CNTL FSS for the 2000 UTC-initialized case

---

[6] Root-mean-square innovations that are comparable to the total ensemble spread (the sum of the assumed standard deviation of the observation error and the ensemble standard deviation) indicate a well-calibrated ensemble.
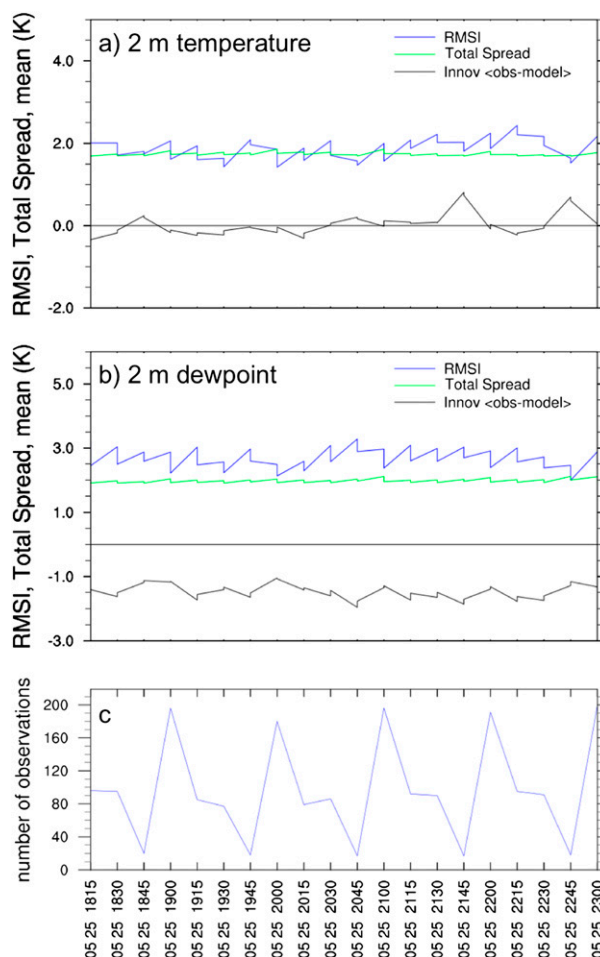

FIG. 14. Ensemble mean innovations (observation minus forecast), root-mean-squared innovation (RMSI), and total ensemble spread (observation standard deviation plus the ensemble standard deviation) evaluated for the METAR 2-m temperature and dewpoint observations for the 25 May 2017 experiment in which the assimilation period was extended to 2300 UTC. The number of observations assimilated at each time is shown in (c). The evaluation area was restricted to a 10° latitude by 12° longitude area centered on CLAMPS.

($\sim$0.65–0.70 for the former and $\sim$0.55–0.65 for the latter). These results give confidence that the AERI and DL retrievals can continue to provide improved mesoscale initial conditions that result in improved near-storm-scale forecasts of convection even when storms are well established in the initial condition and when the baseline skill is already quite high within $\sim$25 km by 25 km neighborhoods. This result could be applicable to efforts to improve the mesoscale background environment for the developmental Warn-on-Forecast system (Jones et al. 2018; Lawson et al. 2018) and for current operational convection-allowing models (Benjamin et al. 2016). These results justify efforts to explore operational
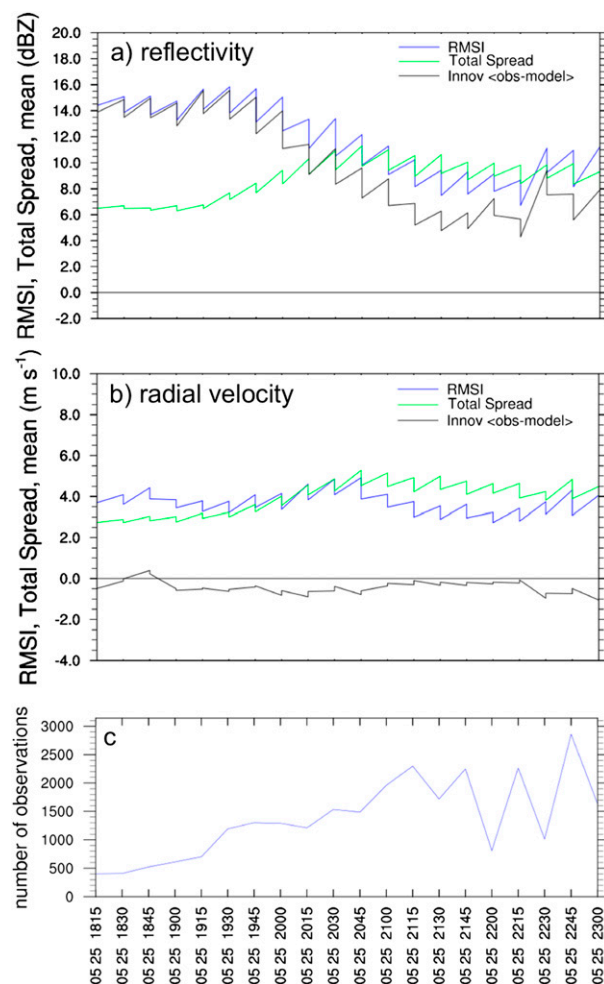
FIG. 15. As in Fig. 14, but for reflectivity and radial velocity observations. The evaluation area was restricted to a 2.5° latitude by 3° longitude area centered on CLAMPS.
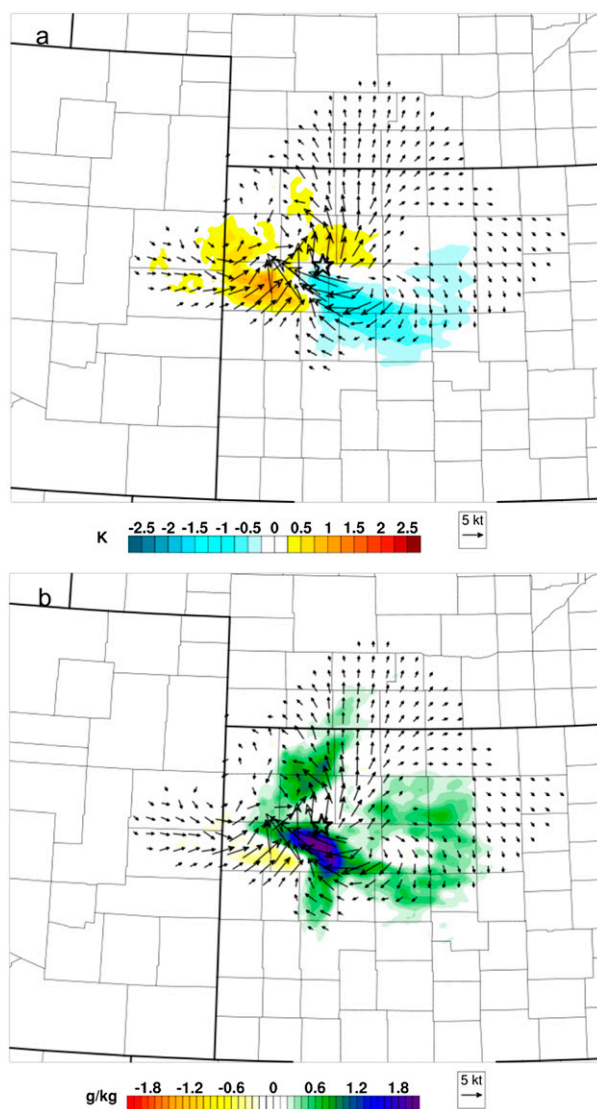

FIG. 16. As in Figs. 7e and 7f, but for the 25 May 2017 experiment in which the assimilation period was extended to 2300 UTC (the valid time of the figures).

implementation of observing systems that can improve the mesoscale background analysis for convection-permitting ensemble forecast systems.

7. Summary and conclusions

In 2016 and 2017, the mobile ground-based profiling system operated by NSSL (CLAMPS) was deployed in the preconvective and near-storm environments in the U.S. Great Plains. Retrievals of temperature and water vapor mixing ratio from an Atmospheric Emitted Radiance Interferometer (AERI) and profiles of horizontal wind components retrieved from a Doppler lidar (DL) were obtained at locations guided by ensemble sensitivity analysis. The goal is to examine the impact of assimilating these profiles on short-term (1–6 h) forecasts of the initiation and early evolution of convection produced by a cycled ensemble analysis and forecasting system. Observing systems like these that can profile the above-ground state in the lowest few kilometers every 2–3 min could be especially important for convective weather applications since the current upper-air radiosonde network is too coarse in space and time to capture mesoscale details than can be important for convective evolution. Furthermore, retrievals of the boundary layer state from satellites is currently, and will remain in the near future, insufficient to capture details of the CBL (e.g., profiles of the vertical wind shear and moisture depth) that also can be important for convective weather forecasting.
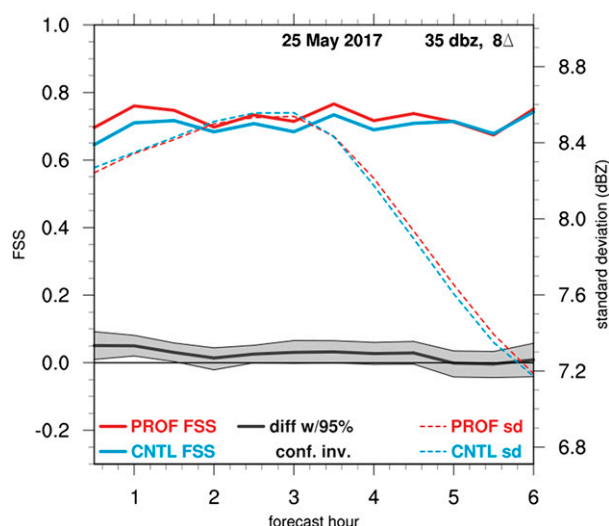
FIG. 17. As in Fig. 9, but for the forecasts for the 25 May 2017 experiment initialized at 2300 UTC.

The analysis and forecast system used here was based on and initialized from the real-time NCAR ensemble. For this study, the analysis was then partially cycled with hourly to 15-min updates on both a meso- and convective-scale domain. Assimilation was performed using the EAKF within WRF-DART and included multiple sources of observations that are routinely assimilated in current operational systems, as well as WSR-88D reflectivity and radial velocity observations, in order to provide a baseline analysis with errors akin to what can be produced operationally. Experiments were performed with assimilating the AERI and DL retrievals together to emulate an observing system that can capture both thermodynamic and kinematic properties of the boundary layer simultaneously (which, in the opinion of the authors, is how such a ground-based profiling system should be proposed as an operational system). The duration of the AERI and DL assimilation varied from 1.5 to 5 h and both analyses and forecasts were produced on a 3-km grid.

The forecast impacts are compared through neighborhood ensemble probabilities (NEPs) of simulated reflectivity ≥ 35 dBZ in ~ 25 km by 25 km neighborhoods to analyses of observed reflectivity. Comparisons of the fractions skill score (FSS), an objective metric that uses the NEPs, indicate relatively few instances of high statistical confidence in the changes made to the FSS by assimilating the AERI and DL retrievals. However, there are far more times when the FSS increases than when the FSS decreases in the 1–6-h forecast period by assimilating the AERI and DL retrievals. The time of maximum forecast impact and the durations of the positive impacts are variable across the forecast period. However, the potential to impact forecasts is best seen

around 3 h, at which time 10 out of the 12 forecasts show an increase in FSS. These results are not sensitive to the reflectivity threshold or the neighborhood size.

Targeting locations for CLAMPS were guided by real-time ensemble sensitivity analysis, similar to how it was applied in MPEX (Weisman et al. 2015) but on much shorter time scales. The hope was that this guidance would increase the likelihood that CLAMPS would sample in an area in which the later forecasts were sensitive to the state of the CBL, which is especially important when evaluating impacts from observations taken at only one location. However, this approach also served as a means to evaluate the ability of ESA fields to explain actual impacts made to forecasts for convective weather applications. To that end, it is difficult to find any consistent relationships between the actual differences in ensemble spread and skill that result from assimilating the observations to both the observation impact values and the forecast sensitivity to individual state variables in the CBL. Many possible reasons for this lack of consistent relationships are discussed above. One possible exception is the forecast sensitivity to the potential temperature averaged in the lowest 1 km AGL. The change in ensemble mean reflectivity that resulted from the assimilation of the AERI and DL retrievals was predicted correctly by the potential temperature sensitivity fields in 11 out of 12 cases.

This study provides evidence that forecasts can be impacted positively by AERI and DL retrievals, even when the retrievals are assimilated through limited cycles (as few as six 15-min cycles) and from only one system at one location. This result is encouraging because the experiments tackled the particularly challenging problem of convective initiation and early convective evolution. An operational system of ground-based profilers would likely not be mobile like the one used here, and would not be limited to one system, but rather would likely be composed of a network of fixed systems, similar to that deployed at the ARM-SGP site. The impacts of retrievals from this site (and from the temporary network of research AERIs and DLs used for PECAN; Geerts et al. 2017) on forecasts of convection are currently being investigated at NSSL and other institutions to help determine if the skill improvements, and the confidence in the skill improvements, would increase if retrievals from this network of profilers are assimilated versus from one targeted AERI/DL system. Furthermore, an operational analysis and forecast system would likely be introducing the AERI and DL retrievals through continuous or partial cycles, in which data is assimilated throughout the morning rather than being introduced at 1800–2000 UTC as in the experiments performed here (tests did show that a longer assimilation

period did translate into improvements in skill). Despite these rather large impediments, the results still provided evidence that the AERI and DL retrievals can positively impact forecasts and warrants continued exploration of the AERI and DL as a means to more frequently observe the above-ground conditions in the lowest few kilometers of the atmosphere.

REFERENCES

Ancell, B., and G. J. Hakim, 2007: Comparing adjoint-and ensemble-sensitivity analysis with applications to observation targeting. *Mon. Wea. Rev.*, **135**, 4117–4134, https://doi.org/10.1175/2007MWR1904.1.

Anderson, J., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642, https://doi.org/10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Avellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Barker, D., and Coauthors, 2012: The Weather Research And Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843, https://doi.org/10.1175/BAMS-D-11-00167.1.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Burlingame, B. M., C. Evans, and P. J. Roebber, 2017: The influence of PBL parameterization on the practical predictability of convection initiation during the Mesoscale Predictability Experiment (MPEX). *Wea. Forecasting*, **32**, 1161–1183, https://doi.org/10.1175/WAF-D-16-0174.1.

Byteway, J. L., and C. D. Kummerow, 2015: Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central U.S. *J. Adv. Model. Earth Syst.*, **7**, 1248–1264, https://doi.org/10.1002/2015MS000497.

Clough, S. A., and M. J. Iacono, 1995: Line-by-line calculation of atmospheric fluxes and cooling rates: 2. Application to carbon dioxide, ozone, methane, nitrous oxide and the halocarbons. *J. Geophys. Res.*, **100**, 16 519–16 535, https://doi.org/10.1029/95JD01386.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather environments. *Wea. Forecasting*, **30**, 591–612, https://doi.org/10.1175/WAF-D-14-00105.1.

——, ——, ——, ——, and I. L. Jirak, 2017: Evaluation of multiple planetary boundary layer parameterization schemes in southeast U.S. cold season severe thunderstorm environments. *Wea. Forecasting*, **32**, 1857–1884, https://doi.org/10.1175/WAF-D-16-0193.1.

Coniglio, M. C., J. Correia Jr., P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF Model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, https://doi.org/10.1175/WAF-D-12-00103.1.

——, S. M. Hitchcock, and K. H. Knopfmeier, 2016: Impact of assimilating preconvective upsonde observations on short-term forecasts of convection observed during MPEX. *Mon. Wea. Rev.*, **144**, 4301–4325, https://doi.org/10.1175/MWR-D-16-0091.1.

Geerts, B., and Coauthors, 2017: The 2015 Plains Elevated Convection at Night Field Project. *Bull. Amer. Meteor. Soc.*, **98**, 767–786, https://doi.org/10.1175/BAMS-D-15-00257.1.

——, and Coauthors, 2018: Recommendations for in situ and remote sensing capabilities in atmospheric convection and turbulence. *Bull. Amer. Meteor. Soc.*, **99**, 2463–2470, https://doi.org/10.1175/BAMS-D-17-0310.1.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2.

Hill, A. J., C. C. Weiss, and B. C. Ancell, 2016: Ensemble sensitivity analysis for mesoscale forecasts of dryline convection initiation. *Mon. Wea. Rev.*, **144**, 4161–4182, https://doi.org/10.1175/MWR-D-15-0338.1.

Hitchcock, S. M., M. C. Coniglio, and K. H. Knopfmeier, 2016: Impact of MPEX upsonde observations on ensemble analyses and forecasts of the 31 May 2013 convective event over Oklahoma. *Mon. Wea. Rev.*, **144**, 2889–2913, https://doi.org/10.1175/MWR-D-15-0344.1.

Jones, T. A., X. Wang, P. Skinner, A. Johnson, and Y. Wang, 2018: Assimilation of *GOES-13* imager clear-sky water vapor (6.5 μm) radiances into a Warn-on-Forecast system. *Mon. Wea. Rev.*, **146**, 1077–1107, https://doi.org/10.1175/MWR-D-17-0280.1.

Knuteson, R., and Coauthors, 2004: Atmospheric Emitted Radiance Interferometer. Part I: Instrument design. *J. Atmos. Oceanic Technol.*, **21**, 1763–1776, https://doi.org/10.1175/JTECH-1662.1.

Lawson, J. R., J. S. Kain, N. Yussouf, D. C. Dowell, D. M. Wheatley, K. H. Knopfmeier, and T. A. Jones, 2018: Advancing from convection-allowing NWP to Warn-on-Forecast: Evidence of progress. *Wea. Forecasting*, **33**, 599–607, https://doi.org/10.1175/WAF-D-17-0145.1.

Lilly, D. K., 1990: Numerical prediction of thunderstorms—Has its time come? *Quart. J. Roy. Meteor. Soc.*, **116**, 779–798, https://doi.org/10.1002/qj.49711649402.

Mittermaier, M., and N. Roberts, 2010: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales

using the fractions skill score. *Wea. Forecasting*, **25**, 343–354, https://doi.org/10.1175/2009WAF2222260.1.

National Research Council, 2009: *Observing Weather and Climate from the Ground up: A Nationwide Network of Networks.* National Academies Press, 250 pp., https://doi.org/10.17226/12540.

Pearson, G., F. Davies, and C. Collier, 2009: An analysis of the performance of the UFAM pulsed Doppler lidar for observing the boundary layer. *J. Atmos. Oceanic Technol.*, **26**, 240–250, https://doi.org/10.1175/2008JTECHA1128.1.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Romine, G. S., C. S. Schwartz, C. Snyder, J. L. Anderson, and M. L. Weisman, 2013: Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Wea. Rev.*, **141**, 1263–1284, https://doi.org/10.1175/MWR-D-12-00112.1.

——, ——, R. D. Torn, and M. L. Weisman, 2016: Impact of assimilating dropsonde observations from mpex on ensemble forecasts of severe weather events. *Mon. Wea. Rev.*, **144**, 3799–3823, https://doi.org/10.1175/MWR-D-15-0407.1.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, https://doi.org/10.1175/2009WAF2222267.1.

——, G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2015: NCAR's experimental real-time convection-allowing ensemble prediction system. *Wea. Forecasting*, **30**, 1645–1654, https://doi.org/10.1175/WAF-D-15-0103.1.

Shin, H. H., and S.-Y. Hong, 2015: Representation of the subgrid-scale turbulent transport in convective boundary layers at gray-zone resolutions. *Mon. Wea. Rev.*, **143**, 250–271, https://doi.org/10.1175/MWR-D-14-00116.1.

——, and J. Dudhia, 2016: Evaluation of PBL parameterizations in WRF at subkilometer grid spacings: Turbulence statistics in the dry convective boundary layer. *Mon. Wea. Rev.*, **144**, 1161–1177, https://doi.org/10.1175/MWR-D-15-0208.1.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032, https://doi.org/10.1175/MWR2830.1.

Sobash, R. A., and D. J. Stensrud, 2013: The impact of covariance localization for radar data on EnKF analyses of a developing MCS: Observing system simulation experiments. *Mon. Wea. Rev.*, **141**, 3691–3709, https://doi.org/10.1175/MWR-D-12-00203.1.

Stalker, J., and Coauthors, 2013: A nationwide network of networks. *Bull. Amer. Meteor. Soc.*, **94**, 1602–1606, https://doi.org/10.1175/1520-0477-94.10.1602.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1500, https://doi.org/10.1175/2009BAMS2795.1.

Torn, R. D., and G. J. Hakim, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663–677, https://doi.org/10.1175/2007MWR2132.1.

——, and G. S. Romine, 2015: Sensitivity of central Oklahoma convection forecasts to upstream potential vorticity anomalies during two strongly forced cases during MPEX. *Mon. Wea. Rev.*, **143**, 4064–4087, https://doi.org/10.1175/MWR-D-15-0085.1.

——, G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490–2502, https://doi.org/10.1175/MWR3187.1.

Turner, D., and U. Löhnert, 2014: Information content and uncertainties in thermodynamic profiles and liquid cloud properties retrieved from the ground-based Atmospheric Emitted Radiance Interferometer (AERI). *J. Appl. Meteor. Climatol.*, **53**, 752–771, https://doi.org/10.1175/JAMC-D-13-0126.1.

——, and G. Blumberg, 2018: Improvements to the AERIoe thermodynamic profile retrieval algorithm. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, https://doi.org/10.1109/JSTARS.2018.2874968, in press.

——, R. Knuteson, H. Revercomb, C. Lo, and R. Dedecker, 2006: Noise reduction of Atmospheric Emitted Radiance Interferometer (AERI) observations using principal component analysis. *J. Atmos. Oceanic Technol.*, **23**, 1223–1238, https://doi.org/10.1175/JTECH1906.1.

Wagner, T. J., P. M. Klein, and D. M. Turner, 2019: A new generation of ground-based mobile platforms for active and passive profiling of the boundary layer. *Bull. Amer. Meteor. Soc.*, **100**, 137–153, https://doi.org/10.1175/BAMS-D-17-0165.1.

Weisman, M. L., and Coauthors, 2015: The Mesoscale Predictability Experiment (MPEX). *Bull. Amer. Meteor. Soc.*, **96**, 2127–2149, https://doi.org/10.1175/BAMS-D-13-00281.1.

Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the nssl experimental warn-on-forecast system. Part I: Radar data experiments. *Wea. Forecasting*, **30**, 1795–1817, https://doi.org/10.1175/WAF-D-15-0043.1.

Wulfmeyer, V., and Coauthors, 2018: A new research approach for observing and characterizing land–atmosphere feedback. *Bull. Amer. Meteor. Soc.*, **99**, 1639–1667, https://doi.org/10.1175/BAMS-D-17-0009.1.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhou, B., M. Xue, and K. Zhu, 2018: A grid-refinement-based approach for modeling the convective boundary layer in the gray zone: Algorithm implementation and testing. *J. Atmos. Sci.*, **75**, 1143–1161, https://doi.org/10.1175/JAS-D-17-0346.1.